Regresión lineal y series de tiempo Gerardo

Sean $(x_1, y_1), ..., (x_n, y_n)$ n parejas de números reales y

$$\bar{x_n} := n^{-1} \sum_{i=1}^{n} x_i \quad y \quad \bar{y_n} := n^{-1} \sum_{i=1}^{n} y_i.$$

Demostrar que

$$\frac{\sum_{i=1}^{n} y_i x_i - n\bar{x_n}\bar{y_n}}{\sum_{i=1}^{n} x_i^2 - n\bar{x_n}^2} = \frac{\sum_{i=1}^{n}(y_i - \bar{y_n})(x_i - \bar{x_n})}{\sum_{i=1}^{n}(x_i - \bar{x_n})^2}.$$

Demostración.

$$\frac{\sum_{i=1}^{n} y_i x_i - n\bar{x_n}\bar{y_n}}{\sum_{i=1}^{n} x_i^2 - n\bar{x_n}^2} = \frac{\sum_{i=1}^{n} y_i x_i - \bar{y_n}\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i^2 - n\bar{x_n}^2}, \tag{1}$$

observemos que como $\sum_{i=1}^{n}(y_i - \bar{y})\bar{x_n} = \bar{x_n}(\sum_{i=1}^{n} y_i - n\bar{y_i}) = 0$, podemos agregar dicha cantidad al numerador de (1), el cual queda de la siguiente manera

$$\begin{aligned}
\sum_{i=1}^{n} y_i x_i - \bar{y_n}\sum x_i &= \sum_{i=1}^{n} y_i x_i - \bar{y_n} x_i - \bar{x_n}\sum(y_i - \bar{y}) \\
&= \sum_{i=1}^{n}(y_i - \bar{y_n})x_i - \bar{x_n}\sum_{i=1}^{n}(y_i - \bar{y_n}) \\
&= \sum_{i=1}^{n}(y_i - \bar{y_n})x_i - \bar{x_n}(y_i - \bar{y_n}) \\
&= \sum_{i=1}^{n}(y_i - \bar{y_n})(x_i - \bar{x_n}). \qquad (2)
\end{aligned}$$

Por otra parte, agregando $n\bar{x_i}^2 - n\bar{x_i}^2 = 0$, en el denominador de (1), tenemos

$$\begin{aligned}\sum_{i=1}^{n} x_i^2 - n\bar{x_i}^2 &= \sum_{i=1}^{n} x_i^2 - 2n\bar{x_i}^2 + n\bar{x_i}^2 \\ &= \sum_{i=1}^{n} x_i^2 - 2\bar{x_i}\sum_{i=1}^{n} x_i + n\bar{x_i}^2 \\ &= \sum_{i=1}^{n} x_i^2 - 2\bar{x_i}x_i + \bar{x_i}^2 \\ &= \sum_{i=1}^{n}(x_i - \bar{x})^2. \end{aligned} \tag{3}$$

Por (1), (2) y (3), concluimos que

$$\frac{\sum_{i=1}^{n} y_i x_i - n\bar{x_n}\bar{y_n}}{\sum_{i=1}^{n} x_i^2 - n\bar{x_n}^2} = \frac{\sum_{i=1}^{n}(y_i - \bar{y_n})(x_i - \bar{x_n})}{\sum_{i=1}^{n}(x_i - \bar{x_n})^2}.$$